

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 21:55:20 ; Search time 1732.53 Seconds
(without alignments)
3084.965 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctcccgcc.....ttgtgaccagggaagagg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	39.6	10.0	418	9	AI973196	AI973196 wt52h09.x
C 2	39	9.8	611	12	B75463	B75463 RPC111-16G3
C 3	38.8	9.8	332	9	BE147008	BE147008 QV4-HT022
C 4	38.8	9.8	333	9	BE147013	BE147013 QV4-HT022
C 5	38.8	9.8	333	9	BE147015	BE147015 QV4-HT022
C 6	38.8	9.8	336	9	BE146800	BE146800 QV4-HT022
C 7	38.4	9.7	875	12	AZ534070	AZ534070 ENTBY52TF
C 8	38.4	9.7	929	12	AZ530962	AZ530962 ENTCH74TF
C 9	38.2	9.6	394	9	BB741171	BB741171 BE741171
C 10	37.8	9.5	495	10	BF392688	BF392688 UI-R-CA0-
C 11	37.6	9.5	340	9	BE146756	BE146756 QV4-HT022
C 12	37.6	9.5	341	9	BE147023	BE147023 QV4-HT022
C 13	37.6	9.5	342	9	BE146946	BE146946 QV4-HT022
C 14	37.6	9.5	702	9	AV722572	AV722572 AV722572
C 15	37.2	9.4	330	9	BE146803	BE146803 QV4-HT022
C 16	37.2	9.4	596	12	AZ974170	AZ974170 2M0248H08
C 17	36.8	9.3	342	9	AA723441	AA723441 zg69b05.s

C 18	36.8	9.3	710	12	BH607591	BH607591 BOKT52TF
C 19	36.6	9.2	1085	12	CNS016YR	AL107373 Drosophil
C 20	36.4	9.2	521	9	AU146546	AU146546 AU146546
C 21	36.4	9.2	720	12	CNS015GI	AL105420 Drosophil
C 22	36.4	9.2	752	12	AZ614340	AZ614340 IM0443B11
C 23	36.4	9.2	942	9	AI519221	AL519221 AL519221
C 24	36.2	9.1	6146	12	AQ839831	AQ839831 260L13-C4
C 25	36	9.1	247	9	BB189866	BB189866 BB189866
C 26	36	9.1	337	9	BE147009	BE147009 QV4-HT022
C 27	35.8	9.0	207	12	AZ782914	AZ782914 2M0024K12
C 28	35.8	9.0	410	12	AZ115236	AZ115236 RPCI-23-1
C 29	35.6	9.0	450	9	AA526521	AA526521 ni60f10.s
C 30	35.6	9.0	604	9	AW559952	AW559952 EST315000
C 31	35.6	9.0	859	12	CNS004YY	AL055406 Drosophil
C 32	35.6	9.0	938	12	BH136715	BH136715 ENT0F89TR
C 33	35.4	8.9	298	9	BE147059	BE147059 QV4-HT022
C 34	35.4	8.9	404	10	BF297102	BF297102 047PDB05
C 35	35.4	8.9	544	12	AZ092469	AZ092469 RPCI-23-4
C 36	35.4	8.9	1086	12	CNS00YXK	AL096962 Drosophil
C 37	35.4	8.9	1204	12	CNS016E2	AL106628 Drosophil
C 38	35.2	8.9	235	9	AI480781	AI480781 ve47g08.x
C 39	35.2	8.9	308	9	AA240821	AA240821 mw24b07.r
C 40	35.2	8.9	515	12	AQ355100	AQ355100 CITBI-El-
C 41	35.2	8.9	624	12	CNS01E33	AL140032 Anopheles
C 42	35.2	8.9	704	12	AQ355013	AQ355013 CITBI-El-
C 43	35	8.8	452	10	H96455	H96455 yt98a03.r1
C 44	35	8.8	491	9	AV724923	AV724923 AV724923
C 45	35	8.8	590	12	AQ763239	AQ763239 HS_3222_B

ALIGNMENTS

RESULT 1
AI973196/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AI973196
wt52h09.x1 NCI_CGAP_Utl1 Homo sapiens CDNA clone IMAGE:2491361 3',
mRNA sequence.
AI973196
AI973196.1 GI:5770022
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 418)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2291 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2491361"
/clone_lib="NCI_CGAP_Utl1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT 109 a 72 c 91 g 146 t
ORIGIN

Query Match 10.0%; Score 39.6; DB 9; Length 418;
Best Local Similarity 53.2%; Pred. No. 1.7;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 39 tctgaataacaatagatgtgctaagatggtatgttcttcggtggtattcttcttggttgcaag 98

Db 159 TCCAAACTCAAGGCATCTACTAAATATATATTATCTCTCTTGATATACCAATCCCGC 100

QY 99 caccaggcatgagtttgggtgaattaatcaagagagtaattcggttacaagaacttgggg 158

Db 99 AACTTTAACTATTATTATGCAGAACTAATCTACATATTAAATGTGCCAATAAATGTTGT 40

QY 159 ttctgtgcaacccaagaaaagtctgaaaataaccaa 196

Db 39 CTTCTCTGTTCCATAAAAAAAAAAAAAAAAAAAAAA 2

RESULT 2

B75463/c
LOCUS B75463 611 bp DNA linear GSS 08-APR-1999
DEFINITION RPCI11-16G3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-16G3, DNA sequence.

ACCESSION B75463 GI:2771150

VERSION GSS.

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 611)

AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter

,J.C.

TITLE Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT Other_GSSs: RPCI11-16G3.TP

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

FEATURES Location/Qualifiers

1..611

/organism="Homo sapiens"

/db_xref="GDB:7505906"

/db_xref="taxon:9606"

/clone="RPCI-11-16G3"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 164 a 139 c 126 g 182 t

ORIGIN

Query Match 9.8%; Score 39; DB 12; Length 611;

Best Local Similarity 51.4%; Pred. No. 2.7;
Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 65 tggatgttcggtgtgtattcttcttggttgcaagcaccaggcatggttttggtaattt 124

Db 541 TGGGATGGCAGACCATAGACTTCATGGTGAAACTACCAAATATGACTTAGCTA 482

QY 125 aatcaagagagtaattcggttacaagaacttggggttctgtcaacccaagaaagtctg 184

Db 481 AGTAAAAAATTCATGTATTGAAATATCTTGGTTCGCTTTAAGAATGAACAAAGTTCT 422

QY 185 aaataaccaaatttagaaaatgtaaggaatcagggggtttcaagactctagata 239

Db 421 AGAGAAACAAGTTTCAGAAATGAACAGTAGTCAAGGTAGCTAGGTGGCTGGCTA 367

RESULT 3

BE147008/c

LOCUS BE147008 332 bp mRNA linear EST 21-JUN-2000

DEFINITION QV4-HT0222-011199-019-d12 HT0222 Homo sapiens cDNA, mRNA sequence.

ACCESSION BE147008

VERSION BE147008.1 GI:8609732

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 332)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-HT0222-011

199-019-d12st3-1999-11-01st4-1)

Seq primer: puc 18 forward

High quality sequence start: 9

High quality sequence stop: 332.

FEATURES Location/Qualifiers

1..332

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0222"

/dev_stage="Adult"

/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 99 a 74 c 93 g 66 t

ORIGIN

Query Match 9.8%; Score 38.8; DB 9; Length 332;

Best Local Similarity 51.8%; Pred. No. 2.8;

Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Db 237 AGCTAATTTTGTATTGTAGTAGAGAGGGGTTTCACCATTTTGGCTAGGCTGATCTTG 178
QY 251 aaaatcctgtctccagggaacctcctcaggggagaatcagctccagctggtttatatcca 310
Db 177 AACTCCTGGCCTCAAGCGATCCACCCGCCCTTGCCTTCCAAAGCTCTGGGTTTACAGGCA 118
QY 311 agtgccattctccccaacct 330
Db 117 TGAGCCACTGTGCCAGCCT 98

RESULT 14
AV722572/c
LOCUS AV722572 702 bp mRNA linear EST 16-OCT-2000
DEFINITION AV722572 HTB Homo sapiens cDNA clone HTBANE07 5', mRNA sequence.
ACCESSION AV722572
VERSION AV722572.1 GI:10825195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)

AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.

TITLE Homo sapiens cDNA HTB clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBANE07"
/clone_lib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 149 a 184 c 154 g 215 t
ORIGIN

Query Match 9.5%; Score 37.6; DB 9; Length 702;
Best Local Similarity 50.6%; Pred. No. 6.7;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 117 gtgaatttaatacaagagagtaattcgttacaagaacttggtgtttctgtgcaacccaaga 176
Db 312 GTGAAATTAAAGGAGATCATGCTTCTTACCAGGAAGTGGGGAATTCAGTTGACCCGTGA 253
QY 177 aaagtctgaaataaaccaaattagaaaatgtaaggaatcaggggtttcaagactctag 236
Db 252 AAAAAGTGAATAAATTAACATGGGGAAGGCTAATTACCATGCCAGTCTGGGGTGGAG 193
QY 237 atagcaagaagtggaaaaatcctgtctccagggaacctcctcaggggagaaatcagctccag 296
Db 192 AACACAGATGGCCAAGGTTGTAGCAGCCTGGGCACCTGCTGAGGCCCATTCAGCGGGAG 133

RESULT 15
BE146803/c
LOCUS BE146803 330 bp mRNA linear EST 21-JUN-2000

DEFINITION QV4-HT0222-181099-013-d07 HT0222 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE146803
VERSION BE146803.1 GI:8609527
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV4-HT0222-181
099-013-d07&t3=1999-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 330.

FEATURES
source
1..330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0222"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 98 a 75 c 95 g 62 t
ORIGIN

Query Match 9.4%; Score 37.2; DB 9; Length 330;
Best Local Similarity 51.2%; Pred. No. 7.7;
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 191 accaaatttagaaaaatgtaaggaatcaaggggttttcaagactctagatagcaagaagtgg 250
Db 225 AGCTAATTTTGTATTGTAGTGGAGAGGGGGTTTCACCATTTGGCTAGGCTGATCTTG 166
QY 251 aaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctggtttatatcca 310
Db 165 AACTCCTGGCCTCAAGCGATCCACCGCCTTGGCCTTCCAAAGCTCTGGGTTTACAGGCA 106
QY 311 agtgccattctccccaacctccaggagtgagagtttcttttagtatgtctt 360
Db 105 TGAGCCACTGTGCCAGCCTGTTTTTTTATTGTTTGTGTTTTTGTGTTTTT 56

Search completed: August 26, 2002, 23:33:24
Job time: 5884 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 23:11:43 ; Search time 280.27 Seconds
(without alignments)
3923.452 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctcccgcc.....ttgtgaccagggaagagg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1757194 seqs, 1388416550 residues

Total number of hits satisfying chosen parameters: 3514388

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq2:*
8: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	396	100.0	396	5	US-09-785-276A-32
2	368.4	93.0	427	5	US-09-785-276A-30405
3	368.4	93.0	427	5	US-09-785-276A-39380
4	351.4	88.7	477	5	US-09-785-276A-9201
5	48.4	12.2	549	6	US-10-198-846-7591
6	47	11.9	523	5	US-09-785-276A-8342
7	46.6	11.8	480	6	US-10-198-846-1504
8	46.4	11.7	454	5	US-09-785-276A-515
9	46.4	11.7	504	6	US-10-198-846-1516
10	46.2	11.7	858	6	US-10-198-846-3335
11	45.8	11.6	410	5	US-09-785-276A-3202
12	45.8	11.6	446	5	US-09-785-276A-5029
13	45.6	11.5	407	5	US-09-785-276A-3821
14	45.6	11.5	439	5	US-09-785-276A-2133
15	45.2	11.4	244	5	US-09-785-276A-739
16	45	11.4	447	6	US-10-198-846-1370
17	44.8	11.3	458	5	US-09-785-276A-495
18	44.6	11.3	492	6	US-10-198-846-1987
19	44.4	11.2	409	5	US-09-785-276A-1180
20	44.2	11.2	385	5	US-09-785-276A-318
21	44.2	11.2	407	5	US-09-785-276A-1000
22	44.2	11.2	852	6	US-10-198-846-6037
23	44.2	11.2	868	6	US-10-198-846-5959
24	44	11.1	383	6	US-10-198-846-1760
25	44	11.1	476	6	US-10-198-846-1432

26	43.8	11.1	427	5	US-09-785-276A-8752	Sequence 8752, Ap
27	43.8	11.1	642	7	US-10-027-632-280816	Sequence 280816,
28	43.8	11.1	670	5	US-09-785-276A-1152	Sequence 1152, Ap
29	43.6	11.0	439	5	US-09-785-276A-477	Sequence 477, App
30	43.6	11.0	440	5	US-09-785-276A-4349	Sequence 4349, Ap
31	43.6	11.0	452	6	US-10-198-846-7840	Sequence 7840, Ap
32	43.6	11.0	454	6	US-10-198-846-583	Sequence 583, App
33	43.6	11.0	474	6	US-10-198-846-664	Sequence 664, App
34	43.6	11.0	865	6	US-10-198-846-1856	Sequence 1856, Ap
35	43.6	11.0	876	6	US-10-198-846-4121	Sequence 4121, Ap
36	43.4	11.0	325	5	US-09-785-276A-8310	Sequence 8310, Ap
37	43.4	11.0	441	5	US-09-785-276A-4348	Sequence 4348, Ap
38	43.2	10.9	255	5	US-09-785-276A-5191	Sequence 5191, Ap
39	43.2	10.9	417	5	US-09-785-276A-6255	Sequence 6255, Ap
40	43.2	10.9	464	6	US-10-198-846-1125	Sequence 1125, Ap
41	43	10.9	304	5	US-09-785-276A-8426	Sequence 8426, Ap
42	42.8	10.8	253	5	US-09-785-276A-4006	Sequence 4006, Ap
43	42.8	10.8	363	5	US-09-785-276A-4022	Sequence 4022, Ap
44	42.8	10.8	380	5	US-09-785-276A-4091	Sequence 4091, Ap
45	42.8	10.8	437	6	US-10-198-846-1411	Sequence 1411, Ap

ALIGNMENTS

RESULT 1
US-09-785-276A-32
; Sequence 32, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF HUMAN PROSTATE CANCER
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-785-276A-32

Query Match	100.0%;	Score 396;	DB 5;	Length 396;
Best Local Similarity	100.0%;	Pred. No. 1.8e-100;		
Matches 396;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	ggcgaattggagctcccccggtggtgtgtattctttgttggaagcaccagcattgagtgct	60		
Db 1	ggcgaattggagctcccccggtggtggtgtgtattctttgttggaagcaccagcattgagtgct	60		
QY 61	aagatggtatgttttcggtgtgtattctttgttggaagcaccagcattgagtgct	120		
Db 61	aagatggtatgttttcggtgtgtattctttgttggaagcaccagcattgagtgct	120		
QY 121	atttaatacgaagagagtaattcgtttacagaacttgggtttctgtgcaaccacgaagaaag	180		

```
Db 121 attaatcaagagagtaattcgttacaagaacttggggtttctgtgcaacccaagaaaag 180
QY 181 tctgaaaaataacaaatttagaaaaatgtaaggaatcagggggtttcagaactctagatag 240
Db 181 tctgaaaaataacaaatttagaaaaatgtaaggaatcagggggtttcagaactctagatag 240
QY 241 caagaagtggaaaaatcctgtctccagggaacctctcctcaggggagagaatcagctccagctgg 300
Db 241 caagaagtggaaaaatcctgtctccagggaacctctcctcaggggagagaatcagctccagctgg 300
QY 301 tttatatccaagtgccattctccccaaacctcaggagtgagagtttcttagtatgtctt 360
Db 301 tttatatccaagtgccattctccccaaacctcaggagtgagagtttcttagtatgtctt 360
QY 361 ggcctctatgccactttgtgaccaggggcaagaggg 396
Db 361 ggcctctatgccactttgtgaccaggggcaagaggg 396

RESULT 2
US-09-785-276A-30405
; Sequence 30405, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30405
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-785-276A-30405

Query Match 93.0%; Score 368.4; DB 5; Length 427;
Best Local Similarity 99.2%; Pred. No. 8.8e-93;
Matches 391; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 4 gaattggagctcc-ccgcggtggcgccgaggtacatctgaattacaatagatgtgctaa 62
Db 1 gaattggagctccaccgcggtggcgccgaggtacatctgaattac-atagatgtgctaa 59
QY 63 gatggtatgttcggtgtgtattcttgggtgcaagcaccaggcatgagtttgggtgaat 122
Db 60 gatggtatgttcggtgtgtattcttgggtgcaagcaccaggcatgagtttgggtgaat 119
QY 123 ttaatcaagagagtaattcgttacagaacttggggttctgtgcaacccaagaaaagtc 182
Db 120 ttaatcaagagagtaattcgttacagaacttggggttctgtgcaacccaagaaaagtc 179
QY 183 tgaataataaccaaatttagaaaaatgtaagggaatcagggggtttcagaactctagatagca 242
Db 180 tgaataataaccaaatttagaaaaatgtaagggaatcagggggtttcagaactctagatagca 239
QY 243 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 302
Db 240 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 299
QY 303 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 362
Db 303 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 362
```

```
QY 243 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 302
Db 240 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 299
QY 303 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 362
Db 300 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 359
QY 363 atcctatgccactttgtgaccaggggcaagagggg 396
Db 360 atcctatgccactttgtgaccaggggcaagagggg 393

RESULT 3
US-09-785-276A-39380
; Sequence 39380, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39380
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-785-276A-39380

Query Match 93.0%; Score 368.4; DB 5; Length 427;
Best Local Similarity 99.2%; Pred. No. 8.8e-93;
Matches 391; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 4 gaattggagctcc-ccgcggtggcgccgaggtacatctgaattacaatagatgtgctaa 62
Db 1 gaattggagctccaccgcggtggcgccgaggtacatctgaattac-atagatgtgctaa 59
QY 63 gatggtatgttcggtgtgtattcttgggtgcaagcaccaggcatgagtttgggtgaat 122
Db 60 gatggtatgttcggtgtgtattcttgggtgcaagcaccaggcatgagtttgggtgaat 119
QY 123 ttaatcaagagagtaattcgttacagaacttggggttctgtgcaacccaagaaaagtc 182
Db 120 ttaatcaagagagtaattcgttacagaacttggggttctgtgcaacccaagaaaagtc 179
QY 183 tgaataataaccaaatttagaaaaatgtaagggaatcagggggtttcagaactctagatagca 242
Db 180 tgaataataaccaaatttagaaaaatgtaagggaatcagggggtttcagaactctagatagca 239
QY 243 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 302
Db 240 agaagtggaaaaatcctgtctccagggaacctcctcaggggagagaatcagctccagctgggtt 299
QY 303 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 362
Db 303 tatatccaagtgccattctccccaaacctccagcaggagtgagagtttcttagtatgtcttgg 362
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 22:32:26 ; Search time 210.62 Seconds
(without alignments)
3228.080 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctcccgcc.....ttgtgaccaggccaagagg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				N_Geneseq_032802:*
1:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1980.DAT:*	
2:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1981.DAT:*	
3:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1982.DAT:*	
4:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1983.DAT:*	
5:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1984.DAT:*	
6:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1985.DAT:*	
7:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1986.DAT:*	
8:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1987.DAT:*	
9:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1988.DAT:*	
10:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1989.DAT:*	
11:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1990.DAT:*	
12:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1991.DAT:*	
13:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1992.DAT:*	
14:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1993.DAT:*	
15:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1994.DAT:*	
16:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1995.DAT:*	
17:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1996.DAT:*	
18:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1997.DAT:*	
19:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1998.DAT:*	
20:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA1999.DAT:*	
21:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA2000.DAT:*	
22:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA2001A.DAT:*	
23:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA2001B.DAT:*	
24:	/SIDS1/gcgdata/	geneseq/	geneseqn-emb1/NA2002.DAT:*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	46.6	11.8	320	22	AAL10155	Human breast cance
2	46.4	11.7	377	22	AAL10167	Human breast cance
3	46.2	11.7	517	22	AAL11997	Human breast cance
4	45	11.4	472	22	AAL08312	Human breast cance
5	44	11.1	247	22	AAL10083	Human breast cance
6	43.6	11.0	447	22	AAL09232	Human breast cance
7	43.6	11.0	555	22	AAL12840	Human breast cance
8	43.2	10.9	423	22	AAL07597	Human breast cance
9	43.2	10.9	449	22	AAL09775	Human breast cance

10	42.8	10.8	128	22	AAL07701	Human breast cance
11	42.8	10.8	247	22	AAL11111	Human breast cance
12	42.8	10.8	257	22	AAL11039	Human breast cance
13	42.8	10.8	312	22	AAL08106	Human breast cance
14	42.8	10.8	403	22	AAL10062	Human breast cance
15	42.8	10.8	1167	22	AAL26699	Human breast cance
16	42.6	10.8	431	22	AAL10341	Human breast cance
17	42.4	10.7	256	22	AAL10126	Human breast cance
18	42.4	10.7	368	22	AAL08243	Human breast cance
19	42	10.6	375	22	AAL10505	Human breast cance
20	42	10.6	427	22	AAL11652	Human breast cance
21	42	10.6	433	22	AAL10669	Human breast cance
22	42	10.6	466	22	AAL09253	Human breast cance
23	42	10.6	469	22	AAL09264	Human breast cance
24	42	10.6	610	22	AAL10486	Human breast cance
25	42	10.6	705	22	AAL10377	Human breast cance
26	41.8	10.6	186	22	AAL10266	Human breast cance
27	41.8	10.6	353	22	AAL11349	Human breast cance
28	41.8	10.6	409	22	AAL08417	Human breast cance
29	41.8	10.6	465	22	AAL11238	Human breast cance
30	41.8	10.6	524	22	AAL07911	Human breast cance
31	41.6	10.5	363	22	AAL10263	Human breast cance
32	41.6	10.5	444	22	AAL07993	Human breast cance
33	41.6	10.5	470	22	AAL09075	Human breast cance
34	41.6	10.5	534	22	AAL07776	Human breast cance
35	41.6	10.5	785	22	AAL12395	Human breast cance
36	41.4	10.5	141	22	AAL07967	Human breast cance
37	41.4	10.5	234	22	AAL10172	Human breast cance
38	41.4	10.5	252	22	AAL09107	Human breast cance
39	41.4	10.5	390	22	AAL09561	Human breast cance
40	41.4	10.5	391	22	AAL10261	Human breast cance
41	41.4	10.5	403	22	AAL09538	Human breast cance
42	41.4	10.5	469	22	AAL10106	Human breast cance
43	41.4	10.5	498	22	AAL09090	Human breast cance
44	41.4	10.5	528	22	AAL11295	Human breast cance
45	41.4	10.5	741	22	AAL11654	Human breast cance

ALIGNMENTS

RESULT		1
AAL10155		
ID	AAL10155 standard; cDNA; 320 BP.	
XX		
AC	AAL10155;	
XX		
DT	07-DEC-2001 (first entry)	
XX		
DE	Human breast cancer expressed polynucleotide 2612.	
XX		
KW	Human; breast cancer; cell marker; cytostatic; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200151628-A2.	
XX		
PD	19-JUL-2001.	
XX		
PF	10-JAN-2001; 2001WO-US00798.	
XX		
PR	14-JAN-2000; 2000US-0176077.	
PR	14-MAR-2000; 2000US-0189167.	
PR	24-MAR-2000; 2000US-0192099.	
PR	29-MAR-2000; 2000US-0193480.	
PR	15-MAY-2000; 2000US-0205230.	
PR	09-JUN-2000; 2000US-0211315.	
PR	25-JUL-2000; 2000US-0220534.	
XX		
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX		
PI	Lillie J, Xu Y, Wang Y, Steinmann K;	
XX		

Db 300 tatatccaagtgcattctcccccaacctccaggagtgagagtttcttttagtatgtctttg 359

QY 363 atcctatgccacactttgtgaccagggcaagagg 396
|||||

Db 360 atcctatgccacactttgtgaccagggcaagagg 393
|||||

RESULT 4

US-09-785-276A-9201

; Sequence 9201, Application US/09785276A

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Endege, Wilson

; APPLICANT: Monahan, John

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF

; TITLE OF INVENTION: HUMAN PROSTATE CANCER

; FILE REFERENCE: MRI-007B

; CURRENT APPLICATION NUMBER: US/09/785,276A

; CURRENT FILING DATE: 2001-02-16

; PRIOR APPLICATION NUMBER: 60/183,319

; PRIOR FILING DATE: 2000-02-17

; PRIOR APPLICATION NUMBER: 60/189,862

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/207,454

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/211,314

; PRIOR FILING DATE: 2000-06-09

; PRIOR APPLICATION NUMBER: 60/219,007

; PRIOR FILING DATE: 2000-07-18

; PRIOR APPLICATION NUMBER: 60/255,281

; PRIOR FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 62232

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 9201

; LENGTH: 477

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-785-276A-9201

Query Match 88.7%; Score 351.4; DB 5; Length 477;

Best Local Similarity 99.5%; Pred. NO. 4.9e-88;

Matches 363; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 32 aggtacatctgaattacaatagatgtgctaagatggttatgttttcggtgtgtattctttg 91
|||||

Db 1 aggtacatctgaattac-atagatgtgctaagatggttatgttcggtgtgtattctttg 59
|||||

QY 92 ttgcaagcaccaggcatgagtttttggtgaatttaataatcaagagagtaattcgtttacaagaa 151
|||||

Db 60 ttgcaagcaccaggcatgagtttttggtgaatttaataatcaagagagtaattcgtttacaagaa 119
|||||

QY 152 cttgggggtttctgtgcaacccaagaaaaagtcgtgaaaaataaccaaatttagaaaaatgtaag 211
|||||

Db 120 cttgggggtttctgtgcaacccaagaaaaagtcgtgaaaaataaccaaatttagaaaaatgtaag 179
|||||

QY 212 gaatcagggggttttcagagactctagatagcaagaagtggaaaaatcctgtctccagggaac 271
|||||

Db 180 gaatcagggggttttcagagactctagatagcaagaagtggaaaaatcctgtctccagggaac 239
|||||

QY 272 ctcctcaggggagaatcagctccagctggttttatatccaagtgccattctcccaacctc 331
|||||

Db 240 ctcctcaggggagaatcagctccagctggttttatatccaagtgccattctcccaacctc 299
|||||

QY 332 caggagtgagagtttcttttagtatgtcttgatcctatgccactttgtgaccagggcaa 391
|||||

Db 300 caggagtgagagtttcttttagtatgtcttgatcctatgccactttgtgaccagggcaa 359
|||||

QY 392 gaggg 396
|||||

Db 360 gaggg 364
|||||

RESULT 5

US-10-198-846-7591

; Sequence 7591, Application US/10198846

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; APPLICANT: Steinmann, Kathleen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MRI-049

; CURRENT APPLICATION NUMBER: US/10/198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 7591

; LENGTH: 549

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 2, 3, 4, 8, 9, 114, 494, 502, 506, 511, 512, 525, 527, 530

; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-7591

Query Match 12.2%; Score 48.4; DB 6; Length 549;

Best Local Similarity 78.4%; Pred. NO. 0.0013;

Matches 58; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ggcgaattggagctccccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
|||||

Db 18 ggcgaattggagctccccgcggtggcgccgaggtacatatatgtgtgtatatatact 77
|||||

QY 61 aagatggtatgttt 74
|||||

Db 78 taaatgtaatatatt 91
|||||

RESULT 6

US-09-785-276A-8342

; Sequence 8342, Application US/09785276A

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Endege, Wilson

; APPLICANT: Monahan, John

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF

; TITLE OF INVENTION: HUMAN PROSTATE CANCER

; FILE REFERENCE: MRI-007B

; CURRENT APPLICATION NUMBER: US/09/785,276A

; CURRENT FILING DATE: 2001-02-16

; PRIOR APPLICATION NUMBER: 60/183,319

; PRIOR FILING DATE: 2000-02-17

; PRIOR APPLICATION NUMBER: 60/189,862

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/207,454

; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/211,314

; PRIOR FILING DATE: 2000-06-09

; PRIOR APPLICATION NUMBER: 60/219,007

; PRIOR FILING DATE: 2000-07-18

; PRIOR APPLICATION NUMBER: 60/255,281

; PRIOR FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 62232

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 8342

; LENGTH: 523

; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4, 328, 442
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-8342

Query Match      11.9%; Score 47; DB 5; Length 523;
Best Local Similarity 74.7%; Pred. No. 0.0031;
Matches 59; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
   |||||
Db 11 ggcgaattggagctcccgcggtggcgccgaggtaccattggtggccaattgattgat 70
   |||||

QY 61 aagatggtatgttttcggtg 79
   |||||
Db 71 ggtaaggaggatcgttg 89

RESULT 7
US-10-198-846-1504
; Sequence 1504, Application US/10198846
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1504
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 99, 101, 109, 118, 121, 124, 180, 181, 189, 203, 204,
; LOCATION: 210, 211, 212, 214, 220, 269, 272, 290, 307, 346, 368, 395,
; LOCATION: 411, 415, 445, 448, 449, 458, 461, 463, 469, 476
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1504

Query Match      11.8%; Score 46.6; DB 6; Length 480;
Best Local Similarity 74.4%; Pred. No. 0.004;
Matches 58; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
   |||||
Db 23 ggcgaattggagctcccgcggtggcgccgaggtaccattggtggccaattgattgat 82
   |||||

QY 61 aagatggtatgttttcggt 78
   |||||
Db 83 ggtaaggaggatcgnt 100

RESULT 8
US-09-785-276A-515
; Sequence 515, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
```

```

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 515
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6, 353, 404, 421, 430
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-515

Query Match      11.7%; Score 46.4; DB 5; Length 454;
Best Local Similarity 70.4%; Pred. No. 0.0045;
Matches 76; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
   |||||
Db 12 ggcgaattggagctcccgcggtggcgccgaggtacttcttaattagaattta-gggct 70
   |||||

QY 61 aagatggtatgttttcggtgtgtattcttggttgcaagcaccaggcat 108
   |||||
Db 71 aacaagagatatataacaataactgtcattatatagaggaggcct 118

RESULT 9
US-10-198-846-1516
; Sequence 1516, Application US/10198846
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1516
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 3, 117, 131, 172, 194, 196, 200, 210, 230, 231, 255, 259,
; LOCATION: 264, 265, 269, 274, 282, 289, 306, 313, 323, 332, 360, 365,
; LOCATION: 376, 381, 404, 411, 416, 419, 437, 444, 451, 452, 459, 468,
; LOCATION: 474, 481, 499
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1516
```


Matches 53; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
|||||
Db 9 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 68
QY 61 aagat 65
|||||
Db 69 aagat 73

RESULT 13
US-09-785-276A-3821
; Sequence 3821, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3821
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 201, 311, 337, 376, 389
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-3821

Query Match 11.5%; Score 45.6; DB 5; Length 407;
Best Local Similarity 75.0%; Pred. No. 0.0072;
Matches 57; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
|||||
Db 9 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 68
QY 61 aagatggtatgttcg 76
|||||
Db 69 aattggattgtgtag 84

RESULT 14
US-09-785-276A-2133
; Sequence 2133, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B

; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2133
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14, 66, 69, 71, 76, 292, 317, 328, 338, 343, 428
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-2133

Query Match 11.5%; Score 45.6; DB 5; Length 439;
Best Local Similarity 74.0%; Pred. No. 0.0074;
Matches 54; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
|||||
Db 5 ggcgaattgnagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 64
QY 61 aagatggtatgtt 73
|||||
Db 65 anagngntttnt 77

RESULT 15
US-09-785-276A-739
; Sequence 739, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 739
; LENGTH: 244
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: 213, 223, 225, 231
; OTHER INFORMATION: n - A,T,C or G
US-09-785-276A-739

Query Match 11.4%; Score 45.2; DB 5; Length 244;
Best Local Similarity 75.7%; Pred. NO. 0.0082;
Matches 56; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattacaatagatgtgct 60
 |||||
Db 1 ggcgaattggagctcccggtggcgccgaggtactcaggcttcaaaaggggtgagga 60
 |||||
QY 61 aagatgggtatgtt 74
 ||| |
Db 61 aaggtcttttgttt 74

Search completed: August 27, 2002, 00:33:20
Job time: 4897 sec

Query Match	11.7%	Score 46.2;	DB 22;	Length 517;
Best Local Similarity	80.6%	Pred. No. 0.00018;		
Matches 54;	Conservative	0;	Mismatches 13;	Indels 0;
				Gaps 0;

Qy 1 ggcgaattggagctcccgcggtggcgccgagggtacatctgaattacaatagatgtct 60
|||||..|||.....|.....|.....|.....|.....|.....|.....|.....|
Db 4 gccgaattggagctcccgcggtggcgccgagggtaccicacaaatttaaaaaaacggqqa 63
|||||..|||.....|.....|.....|.....|.....|.....|.....|.....|

RESULT	4	
AAL08312		
ID	AAL08312	standard; cDNA; 472 BP.
XX		
XX	AC	AAL08312;
XX		
DT	07-DEC-2001	(first entry)
XX		
DE		Human breast cancer expressed polynucleotide 769.
XX		
KW		Human; breast cancer; cell marker; cytostatic; ss.

Query Match 11.4%; Score 45; DB 22; Length 472;
Best Local Similarity 82.3%; Pred. No. 0.00041;
Matches 51; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ggcgaaattggagctccccgcggtggcggccgagggtacatctgaattacaatagatgtgct 60
||| ||||||||||||||||||||||||||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 ggcnaattggagctccccgcggtggcggccgagggtacagcttaattggcaaaagaagaatct 64

Qy 61 aa 62
Pb 65 ta 66

RESULT	5
AAL10083	
ID	AAL10083 standard; cDNA; 247 BP.
XX	
AC	AAL10083;
XX	
DT	07-DEC-2001 (first entry)
XX	
DE	Human breast cancer expressed polynucleotide 2540.
XX	
KW	Human: breast cancer; cell marker; cytostatic; ss.

Query Match	11.1%;	Score 44;	DB 22;	Length 247;
Best Local Similarity	53.5%;	Pred. No. 0.00064;		
Matches 92;	Conservative	0;	Mismatches 80;	Indels 0;
				Gaps 0;

[illegible]

```
RESULT 6
AAL09232
ID AAL09232 standard; cDNA; 447 BP.
XX
AC AAL09232;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 1689.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
WPI; 2001-451856/48.
XX
DR New peptide useful as a marker for the diagnosis of breast cancer
XX
PT
XX
PS Claim 1; Page 338-339; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 447 BP; 117 A; 84 C; 93 G; 153 T; 0 other;

Query Match 11.0%; Score 43.6; DB 22; Length 447;
Best Local Similarity 84.5%; Pred. No. 0.0011;
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattacaatagtg 58
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 ggcgaattggagctcccggtggcgccgaggtacagcatggataatttagtg 69

RESULT 7
AAL12840
ID AAL12840 standard; cDNA; 555 BP.
XX
AC AAL12840;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 5297.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
```

```
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
WPI; 2001-451856/48.
XX
DR New peptide useful as a marker for the diagnosis of breast cancer
XX
PT
XX
PS Claim 1; Page 951; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 555 BP; 130 A; 148 C; 136 G; 141 T; 0 other;

Query Match 11.0%; Score 43.6; DB 22; Length 555;
Best Local Similarity 92.0%; Pred. No. 0.0012;
Matches 46; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattacaa 50
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 ggcgaattggagctcccggtggcgccgaggtacatcttaataagaaa 50

RESULT 8
AAL07597
ID AAL07597 standard; cDNA; 423 BP.
XX
AC AAL07597;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 54.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
```


CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX Sequence 128 BP; 41 A; 23 C; 28 G; 32 T; 4 other;
SQ

Query Match 10.8%; Score 42.8; DB 22; Length 128;
Best Local Similarity 89.8%; Pred. No. 0.0012;
Matches 44; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattaca 49
|||||
Db 9 ggcgaattggagctcccggtggcgccgaggtacatnnnaatnata 57
|||||

RESULT 11
AAL11111
ID AAL11111 standard; cDNA; 247 BP.
XX
AC AAL11111;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 3568.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 657; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 247 BP; 68 A; 69 C; 62 G; 48 T; 0 other;

Query Match 10.8%; Score 42.8; DB 22; Length 247;
Best Local Similarity 87.0%; Pred. No. 0.0015;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattacaataga 54
|||||

Db 4 ggcgaattggagctcccggtggcgccgaggtacaaactgaagcaagagga 57

RESULT 12
AAL11039
ID AAL11039 standard; cDNA; 257 BP.
XX
AC AAL11039;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 3496.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 643-644; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 257 BP; 71 A; 74 C; 62 G; 50 T; 0 other;

Query Match 10.8%; Score 42.8; DB 22; Length 257;
Best Local Similarity 87.0%; Pred. No. 0.0016;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccggtggcgccgaggtacatctgaattacaataga 54
|||||
Db 13 ggcgaattggagctcccggtggcgccgaggtacaaactgaagcaagagga 66
|||||

RESULT 13
AAL08106
ID AAL08106 standard; cDNA; 312 BP.
XX
AC AAL08106;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 563.
XX

KW Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer

XX Claim 1; Page 185; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

XX Sequence 312 BP; 108 A; 63 C; 71 G; 67 T; 3 other;

Query Match 10.8%; Score 42.8; DB 22; Length 312;
Best Local Similarity 74.6%; Pred. No. 0.0017;
Matches 53; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaaatagtgtgct 60
|||||
Db 7 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaaatagtgaag 66

QY 61 aagatggtatg 71

Db 67 aaaattnttg 77

RESULT 14

AAL10062

ID AAL10062 standard; cDNA; 403 BP.

XX AAL10062;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 2519.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer

XX Claim 1; Page 479-480; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterising treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

XX Sequence 403 BP; 106 A; 80 C; 118 G; 96 T; 3 other;

Query Match 10.8%; Score 42.8; DB 22; Length 403;
Best Local Similarity 95.7%; Pred. No. 0.0019;
Matches 44; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaatt 46
|||||

Db 6 ggcgaattggagctcccgcggtggcgccgaggtacatctgaatt 51

RESULT 15

AAL26699/c

ID AAL26699 standard; cDNA; 1167 BP.

XX AAL26699;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 19156.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX WO200151628-A2.

XX 19-JUL-2001.

XX 10-JAN-2001; 2001WO-US00798.

XX 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
XX
PS Claim 1; Page 3645; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 1167 BP; 298 A; 246 C; 295 G; 327 T; 1 other;

Query Match 10.8%; Score 42.8; DB 22; Length 1167;
Best Local Similarity 75.7%; Pred. No. 0.003;
Matches 53; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 ggcgaattggagctcccgcggtggcgccgaggtacatctgaattacaatagatgtgct 60
|||
Db 1164 GSCGAATTGGAGCTCCCGCGGTGCGCGCGAGGTACATGCAGGGTATATTACAATTGCA 1105
|||

QY 61 aagatggtat 70
|
Db 1104 ATCATTTATT 1095

Search completed: August 27, 2002, 00:27:10
Job time: 6884 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2002, 21:26:00 ; Search time 1896.6 Seconds
(without alignments)
4369.350 Million cell updates/sec

Title: US-09-785-276A-32
Perfect score: 396
Sequence: 1 ggcgaattggagctcccgcc.....ttgtgaccaggggcaagaggg 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:**

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

1	362.2	91.5	143800	2	AC011875	AC011875 Homo sapi
2	362.2	91.5	186962	9	AP000848	AP000848 Homo sapi
3	360.6	91.1	115762	2	AP000663	AP000663 Homo sapi
4	348.6	88.0	172830	2	AP001320	AP001320 Homo sapi
5	45.4	11.5	45029	2	AC017280	AC017280 Drosophil
6	45.4	11.5	68626	2	AC006168	AC006168 Drosophil
7	45.4	11.5	102621	3	DMBR43E12	AL138971 Drosophil
8	45.4	11.5	181443	3	AC104148	AC104148 Drosophil
9	45.4	11.5	300205	3	AE003425	AE003425 Drosophil
10	44.6	11.3	161123	9	AC011029	AC011029 Homo sapi
11	44.2	11.2	150152	2	AL355502	AL355502 Homo sapi
12	43	10.9	113559	9	AC025578	AC025578 Homo sapi
13	42.8	10.8	159229	9	AL359180	AL359180 Human DNA
14	42.6	10.8	113254	9	AL137009	AL137009 Human DNA
15	42.4	10.7	259894	9	HUAC002302	AC002302 Homo sapi
16	41.8	10.6	114412	9	AL365260	AL365260 Human DNA
17	41.8	10.6	153046	9	AL590009	AL590009 Human DNA
18	41.8	10.6	157698	2	AC025692	AC025692 Homo sapi
19	41.8	10.6	179463	2	AL596453	AL596453 Homo sapi
20	41.8	10.6	213294	2	AC099778	AC099778 Homo sapi
21	41.4	10.5	154918	9	HSDJ67A8	AL121957 Human DNA
22	41.4	10.5	155362	2	AC022244	AC022244 Homo sapi
23	41.2	10.4	410	6	AX284391	AX284391 Sequence
24	41.2	10.4	166095	9	AC006052	AC006052 Homo sapi
25	41	10.4	171640	30	AC055875	AC055875 Homo sapi
26	41	10.4	186773	2	AC106900	AC106900 Homo sapi
27	40.8	10.3	306	6	AX284580	AX284580 Sequence
28	40.2	10.2	99011	2	AC004937	AC004937 Homo sapi
29	40.2	10.2	110000	9	AC005079_0	AC005079 Homo sapi
30	40.2	10.2	161898	2	AC021683	AC021683 Homo sapi
31	40.2	10.2	164950	2	AC105191	AC105191 Homo sapi
32	40	10.1	182	6	AX284268	AX284268 Sequence
33	40	10.1	71474	9	AC098819	AC098819 Homo sapi
34	40	10.1	161872	2	AC015949	AC015949 Homo sapi
35	39.8	10.1	42418	9	AL445990	AL445990 Human DNA
36	39.8	10.1	186102	2	AC015905	AC015905 Homo sapi
37	39.6	10.0	89837	2	AC091822	AC091822 Homo sapi
38	39.6	10.0	122050	2	AC091512	AC091512 Leishmani
39	39.6	10.0	133715	3	AC087838	AC087838 Leishmani
40	39.6	10.0	204767	2	AC011367	AC011367 Homo sapi
41	39.6	10.0	251664	2	AC008533	AC008533 Homo sapi
42	39.4	9.9	633	6	AX284552	AX284552 Sequence
43	39.4	9.9	127593	9	AC006236	AC006236 Homo sapi
44	39.4	9.9	161920	9	AC025169	AC025169 Homo sapi
45	39.4	9.9	166895	2	AC023415	AC023415 Homo sapi

ALIGNMENTS

RESULT	1	AC011875	143800 bp	DNA	linear	HTG 12-MAR-2000
LOCUS	AC011875	Homo sapiens clone RP11-16K5,	WORKING DRAFT SEQUENCE,	26	unordered	
DEFINITION	AC011875	Homo sapiens clone RP11-16K5,	WORKING DRAFT SEQUENCE,	26	unordered	
ACCESSION	AC011875	pieces.				
VERSION	AC011875.3	GI:7107950				
KEYWORDS	HTG; HTGS_PHASE1;	HTGS_DRAFT.				
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 143800)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.					
TITLE	Homo sapiens, clone RP11-16K5					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 143800)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castile,A., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,					

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Direct Submission
Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6453961.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3566
Center clone name: 16_K_5

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 93387 bases at least Q40
Consensus quality: 115701 bases at least Q30
Consensus quality: 130381 bases at least Q20
Insert size: 141000; agarose-fp
Insert size: 141300; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1029: contig of 1029 bp in length
1030 1129: gap of 100 bp
1130 2964: contig of 1835 bp in length
2965 3064: gap of 100 bp
3065 4253: contig of 1189 bp in length
4254 4353: gap of 100 bp
4354 5678: contig of 1325 bp in length
5679 5778: gap of 100 bp
5779 6855: contig of 1077 bp in length
6856 6955: gap of 100 bp
6956 8847: contig of 1892 bp in length
8848 8947: gap of 100 bp
8948 10945: contig of 1998 bp in length
10946 11045: gap of 100 bp
11046 13001: contig of 1956 bp in length
13002 13101: gap of 100 bp
13102 15430: contig of 2329 bp in length
15431 15530: gap of 100 bp
15531 17919: contig of 2389 bp in length
17920 18019: gap of 100 bp
18020 20135: contig of 2116 bp in length
20136 20235: gap of 100 bp
20236 23750: contig of 3515 bp in length
23751 23850: gap of 100 bp
23851 26794: contig of 2944 bp in length
26795 26894: gap of 100 bp
26895 29482: contig of 2588 bp in length
29483 29582: gap of 100 bp
29583 33174: contig of 3592 bp in length
33175 33274: gap of 100 bp

* 33275 35185: contig of 1911 bp in length
* 35186 35285: gap of 100 bp
* 35286 39745: contig of 4460 bp in length
* 39746 39845: gap of 100 bp
* 39846 44222: contig of 4377 bp in length
* 44223 44322: gap of 100 bp
* 44323 48990: contig of 4668 bp in length
* 48991 49090: gap of 100 bp
* 49091 57790: contig of 8700 bp in length
* 57791 57890: gap of 100 bp
* 57891 66822: contig of 8932 bp in length
* 66823 66922: gap of 100 bp
* 66923 76709: contig of 9787 bp in length
* 76710 76809: gap of 100 bp
* 76810 92865: contig of 16056 bp in length
* 92866 92965: gap of 100 bp
* 92966 106278: contig of 13313 bp in length
* 106279 106378: gap of 100 bp
* 106379 123041: contig of 16663 bp in length
* 123042 123141: gap of 100 bp
* 123142 143800: contig of 20659 bp in length.

FEATURES

Location/Qualifiers

Source

1. .143800
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-16K5"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. .1029
/note="assembly_fragment"

misc_feature

1130. .2964
/note="assembly_fragment"

misc_feature

3065. .4253
/note="assembly_fragment"

misc_feature

4354. .5678
/note="assembly_fragment"

misc_feature

5779. .6855
/note="assembly_fragment"

misc_feature

6956. .8847
/note="assembly_fragment"

misc_feature

8948. .10945
/note="assembly_fragment"

misc_feature

11046. .13001
/note="assembly_fragment"

misc_feature

13102. .15430
/note="assembly_fragment"

misc_feature

15531. .17919
/note="assembly_fragment"

misc_feature

18020. .20135
/note="assembly_fragment"

misc_feature

20236. .23750
/note="assembly_fragment"

misc_feature

23851. .26794
/note="assembly_fragment"

misc_feature

26895. .29482
/note="assembly_fragment"

misc_feature

29583. .33174
/note="assembly_fragment"

misc_feature

33275. .35185
/note="assembly_fragment"

clone_end:SP6
vector_side:right"

misc_feature

35286. .39745
/note="assembly_fragment"

misc_feature

39846. .44222
/note="assembly_fragment"

misc_feature

44323. .48990
/note="assembly_fragment"

misc_feature

49091. .57790
/note="assembly_fragment"

misc_feature

57891. .66822
/note="assembly_fragment"

misc_feature

66923. .76709
/note="assembly_fragment"

misc_feature

6923. .76709
/note="assembly_fragment"

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: CMB9-3F20
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 105451 bases at least Q40
Consensus quality: 110431 bases at least Q30
Consensus quality: 112519 bases at least Q20
Insert size: 113762; sum-of-contigs
Quality coverage: 4.53x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 27996 contig of 27996 bp in length
28097 44759 contig of 16663 bp in length
44860 53606 contig of 8747 bp in length
53707 61270 contig of 7564 bp in length
61371 66307 contig of 4937 bp in length
66408 71338 contig of 4931 bp in length
71439 77284 contig of 5846 bp in length
77385 81759 contig of 4375 bp in length
81860 86481 contig of 4622 bp in length
86582 89025 contig of 2444 bp in length
89126 92534 contig of 3409 bp in length
92635 95988 contig of 3354 bp in length
96089 99742 contig of 3654 bp in length
99843 103044 contig of 3202 bp in length
103145 105701 contig of 2956 bp in length
105802 108757 contig of 2127 bp in length
108858 110984 contig of 1842 bp in length
111085 112926 contig of 1384 bp in length
113027 114410 contig of 53 bp in length
114511 114563 contig of 1099 bp in length
114664 115762 contig of 1099 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 27996: contig of 27996 bp in length
* 27997 28096: gap of 100 bp
* 28097 44759: contig of 16663 bp in length
* 44760 44859: gap of 100 bp
* 44860 53606: contig of 8747 bp in length
* 53607 53706: gap of 100 bp
* 53707 61270: contig of 7564 bp in length
* 61271 61370: gap of 100 bp
* 61371 66307: contig of 4937 bp in length
* 66308 66407: gap of 100 bp
* 66408 71338: contig of 4931 bp in length
* 71339 71438: gap of 100 bp
* 71439 77284: contig of 5846 bp in length
* 77285 77384: gap of 100 bp
* 77385 81759: contig of 4375 bp in length
* 81760 81859: gap of 100 bp
* 81860 86481: contig of 4622 bp in length
* 86482 86581: gap of 100 bp

* 86582 89025: contig of 2444 bp in length
* 89026 89125: gap of 100 bp
* 89126 92534: contig of 3409 bp in length
* 92535 92634: gap of 100 bp
* 92635 95988: contig of 3354 bp in length
* 95989 96088: gap of 100 bp
* 96089 99742: contig of 3654 bp in length
* 99743 99842: gap of 100 bp
* 99843 103044: contig of 3202 bp in length
* 103045 103144: gap of 100 bp
* 103145 105701: contig of 2957 bp in length
* 105702 105801: gap of 100 bp
* 105802 108757: contig of 2956 bp in length
* 108758 108857: gap of 100 bp
* 108858 110984: contig of 2127 bp in length
* 110985 111084: gap of 100 bp
* 111085 112926: contig of 1842 bp in length
* 112927 113026: gap of 100 bp
* 113027 114410: contig of 1384 bp in length
* 114411 114510: gap of 100 bp
* 114511 114563: contig of 53 bp in length
* 114564 114663: gap of 100 bp
* 114664 115762: contig of 1099 bp in length.

FEATURES

source

Location/Qualifiers
1..115762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q21"
/clone="CMB9-3F20"

misc_feature 1..27996
/note="assembly_fragment clone_end:SP6 vector_side:left"

misc_feature 28097..44759
/note="assembly_fragment"

misc_feature 44860..53606
/note="assembly_fragment"

misc_feature 53707..61270
/note="assembly_fragment"

misc_feature 61371..66307
/note="assembly_fragment"

misc_feature 66408..71338
/note="assembly_fragment"

misc_feature 71439..77284
/note="assembly_fragment"

misc_feature 77385..81759
/note="assembly_fragment"

misc_feature 81860..86481
/note="assembly_fragment"

misc_feature 86582..89025
/note="assembly_fragment"

misc_feature 89126..92534
/note="assembly_fragment"

misc_feature 92635..95988
/note="assembly_fragment"

misc_feature 96089..99742
/note="assembly_fragment"

misc_feature 99843..103044
/note="assembly_fragment"

misc_feature 103145..105701
/note="assembly_fragment"

misc_feature 105802..108757
/note="assembly_fragment"

misc_feature 108858..110984
/note="assembly_fragment"

misc_feature 111085..112926
/note="assembly_fragment"

misc_feature 113027..114410
/note="assembly_fragment"

misc_feature 114511..114563
/note="assembly_fragment clone_end:T7 vector_side:right"

misc_feature 114664..115762
/note="assembly_fragment"

BASE COUNT 34122 a 22523 c 22320 g 34797 t 2000 others

ORIGIN

Query Match 91.1%; Score 360.6; DB 2; Length 115762;
Best Local Similarity 98.9%; Pred. No. 3.2e-98;
Matches 363; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 30 cgaggtacatctgaattacaatagatgtgctaagatggtatgtttcgggtgtgtattctttt 89
|| |||||
Db 68526 CGTAGTACATCTGAATTACAATAGATGTGCTAAGATGGTATGTGTGGTGTATCTTTT 68585

QY 90 ggttgaagcaccaggcatgagtttgggtgaatttaatacaagagagtaattcgttacaag 149
|| |||||
Db 68586 GGTGCAAGCACCAGGCATGAGTTTGGTGAATTTAATCAAGAGAGTAATTCGTTACAAG 68645

QY 150 aacttgggggttctgtcaaccacaagaaaagtctgaaaaataaccaaatttagaaaatgta 209
|| |||||
Db 68646 AACTTGGGGTTTCTGTGCAACCCCAAGAAAGTCTGAAAAATAACCAAAATTTAGAAAAATGTA 68705

QY 210 aggaatcagggggtttcaagactctagatagcaagaagtggaaaaatcctgtctccaggga 269
|| |||||
Db 68706 AGGAATCAGGGGGTTTCAAGACTCTAGATAGCAAGAAGTGGAAAAATCCTGTCTCCAGGGA 68765

QY 270 acctcctcaggggagaaatcagctccagctgggtttatatccaaagtccattctcccaacc 329
|| |||||
Db 68766 ACCTCCTCAGGGGAGATCAGCTCCAGCTGGTTTATATCCAAAGTGCCATTCCTCCCAACC 68825

QY 330 tccaggagtgaagtttcttagtatgtctctggatcctatgcccaactttgtgaccagggc 389
|| |||||
Db 68826 TCCAGGAGTGAGAGTTTCTTTAGTATGTCTTGGATCCTATGCCCACTTTGTGACCAGGGC 68885

QY 390 aagaggg 396
|||||
Db 68886 AAGAGGG 68892

RESULT 4

AP001320
LOCUS AP001320 172830 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-79904 map 11q14, WORKING
DRAFT SEQUENCE, 32 unordered pieces.

ACCESSION AP001320
VERSION AP001320.2 GI:8117247
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-79904.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 172830)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 172,830 genomic DNA of 11q14
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 172830)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT On May 30, 2000 this sequence version replaced gi:7209763.

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-79904
----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 145836 bases at least Q40
Consensus quality: 159404 bases at least Q30
Consensus quality: 166388 bases at least Q20
Insert size: 169730; sum-of-contigs
Quality coverage: 4.17x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 15732 contig of 15732 bp in length
15833 34101 contig of 18269 bp in length
34202 47479 contig of 13278 bp in length
47580 57756 contig of 10177 bp in length
57857 69426 contig of 11570 bp in length
69527 80185 contig of 10659 bp in length
80286 88989 contig of 8704 bp in length
89090 99362 contig of 10273 bp in length
99463 103433 contig of 3971 bp in length
103534 109064 contig of 5531 bp in length
109165 113855 contig of 4691 bp in length
113956 119140 contig of 5185 bp in length
119241 124661 contig of 5421 bp in length
124762 126822 contig of 2061 bp in length
126923 131337 contig of 4415 bp in length
131438 135078 contig of 3641 bp in length
135179 138134 contig of 2956 bp in length
138235 142123 contig of 3889 bp in length
142224 144719 contig of 2496 bp in length
144820 149030 contig of 4211 bp in length
149131 151134 contig of 2004 bp in length
151235 153751 contig of 2517 bp in length
153852 156504 contig of 2653 bp in length
156605 159583 contig of 2979 bp in length
159684 161669 contig of 1986 bp in length
161770 164135 contig of 2366 bp in length
164236 165854 contig of 1619 bp in length
165955 167495 contig of 1541 bp in length
167596 168771 contig of 1176 bp in length
168872 170322 contig of 1451 bp in length
170423 171562 contig of 1140 bp in length
171663 172830 contig of 1168 bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 15732: contig of 15732 bp in length
* 15733 15832: gap of 100 bp
* 15833 34101: contig of 18269 bp in length
* 34102 34201: gap of 100 bp
* 34202 47479: contig of 13278 bp in length
* 47480 47579: gap of 100 bp
* 47580 57756: contig of 10177 bp in length
* 57757 57856: gap of 100 bp
* 57857 59426: contig of 11570 bp in length
* 59427 69526: gap of 100 bp
* 69527 80185: contig of 10659 bp in length
* 80186 80285: gap of 100 bp
* 80286 88989: contig of 8704 bp in length
* 88990 89089: gap of 100 bp
* 89090 99362: contig of 10273 bp in length
* 99363 99462: gap of 100 bp

Db 27283 AAATATAATTGAGCTAATGAATAACAATAAGCCGAAATGCTAATGGAGTAGATCAAAA 27342

QY 169 acccaagaaaaagtctgaaaaataacccaaatttagaaaaatgtaaggaaatcagggggtttcaa 228

Db 27343 ATCAAAAAAAGATAAAACAAATTATATATATAAGTAATAATAAGGAATAATGTA 27402

QY 229 gactctagatagcaagaag 247

Db 27403 TATACCAGATGGCGACGAG 27421

RESULT 7

DMBR43E12/c 102621 bp DNA linear INV 07-FEB-2000

LOCUS DMBR43E12 102621 bp DNA linear INV 07-FEB-2000

DEFINITION Drosophila melanogaster BAC clone BACR43E12.

ACCESSION AL138971

VERSION AL138971.1 GI:6940335

KEYWORDS

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 102621)

AUTHORS Ferraz,C., Vidal,S., Brun,C., Bucheton,A. and Demaille,J.G.

TITLE Sequencing the distal X chromosome of Drosophila melanogaster

JOURNAL Unpublished

REMARK Institut de Genetique Humaine -UPR 1142- CNRS, 141,rue de la Cardonille, 34396 Montpellier Cedex 5, France

REFERENCE 2 (bases 1 to 102621)

AUTHORS Benos,P.

TITLE Direct Submission

JOURNAL Submitted (05-FEB-2000) European Drosophila Genome Sequencing Consortium

COMMENT Sequence submitted by Takis Benos, EMBL Outstation - The EBI, Hinxton, Cambridge, CB10 1SD, U.K.

E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome Sequencing Consortium. For further information see the European Drosophila Genome Sequencing Consortium's web site: http://edgp.ebi.ac.uk/.

The syntax for the representation of annotation used in this record is documented at: ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation.README

Coding sequences are predicted from computer analysis, using both gene and CDS prediction programs and matches to other sequences. These predictions and matches have been evaluated by the annotators and may have been refined by hand (in which case a Genefinder prediction will have no score. The annotators have also used their judgement on what matches to represent in this record. A far more complete annotation record is available from FlyBase (http://flybase.bio.indiana.edu/) through the FlyBase Annotation Object linked by the db_xref qualifier in the Feature Table. IMPORTANT: This sequence is NOT necessarily the entire insert of clone BACR43E12. It may be shorter, since we are minimising the overlap between clones to the left with clone BACN33B1 The true left end of clone BACR43E12 is unknown

Clone BACR43E12 overlaps to the right with clone 100G7 The true right end of clone BACR43E12 is unknown

Sequence in absolute orientation with respect to chromosome Clone=BACR43E12; Contig ID=28; Length=102621; Status=Finished.

FEATURES

Location/Qualifiers

1..102621

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BAC BACR43E12"

6318..7103

/gene="EG:BACR43E12.1"

6318..7103

/gene="EG:BACR43E12.1"

/note="/prediction=(method:'genefinder', version:'084', score:'29.13'); prediction=(method:'gensecan', version:'1.0',

42208 42287: gap of unknown length

42288 45740: contig of 3453 bp in length

45741 45820: gap of unknown length

45821 52324: contig of 6504 bp in length

52325 52404: gap of unknown length

52405 53166: contig of 762 bp in length

53167 53246: gap of unknown length

53247 53978: contig of 732 bp in length

53979 54058: gap of unknown length

54059 54639: contig of 581 bp in length

54640 54719: gap of unknown length

54720 55377: contig of 658 bp in length

55378 56074: contig of 617 bp in length

55458 56075 56154: gap of unknown length

56075 56155 56785: contig of 631 bp in length

56786 56865: gap of unknown length

56866 57505: contig of 640 bp in length

57506 57585: gap of unknown length

57586 58418: contig of 833 bp in length

58419 58499 59104: gap of unknown length

58499 59185 59825: contig of 641 bp in length

59185 59826 60665: contig of 760 bp in length

59906 60745: gap of unknown length

60746 61377: contig of 632 bp in length

61378 61457: gap of unknown length

61458 62118: contig of 661 bp in length

62119 62198: gap of unknown length

62199 62791: contig of 593 bp in length

62792 62871: gap of unknown length

62872 63524: contig of 653 bp in length

63525 63604: gap of unknown length

63605 64223: contig of 619 bp in length

64224 64303: gap of unknown length

64304 64965: contig of 662 bp in length

64966 65045: gap of unknown length

65046 65693: contig of 648 bp in length

65694 65773: gap of unknown length

65774 66389: contig of 616 bp in length

66390 66469: gap of unknown length

66470 67127: gap of unknown length

67048 67128 67841: contig of 714 bp in length

67842 67921: gap of unknown length

67922 68626: contig of 705 bp in length.

FEATURES

source

Location/Qualifiers

1..68626

/organism="Drosophila melanogaster"

/strain="Y; cn bw sp"

/db_xref="taxon:7227"

/chromosome="X"

/map="3C-3C"

/clone="BACR48C06 (D485) RPCI-98 48.C.6"

/clone_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6"

BASE COUNT 17635 a 14487 c 13985 g 18195 t 4324 others

ORIGIN

Query Match 11.5%; Score 45.4; DB 2; Length 68626;

Best Local Similarity 51.8%; Pred. No. 0.0085;

Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 49 aatagatgtgctaagatggttatgttttcggtgtgtattcttcttggttgcaagcaccagggcat 108

Db 27223 AACAGAAGGCCCTTTAATTATATATGTGCAGTAGTTTGTGTGAGCTTAAGCAAGCCGTAT 27282

QY 109 gagttttggtgaatttaataagaagagagtaattcgttacagaacttgggggtttctgtgca 168


```
score: '84.99'); /match=(desc: 'MYELOBLAST KIAA0252
(FRAGMENT)', species: 'Homo sapiens (Human)',
ranges: (query: 6441..6860,
target: SPTREMBL: Q92541: 298..437, score: '210.00')),
method: 'blastx', version: '1.4.9')
/codon_start=1
/protein_id="CAB72279.1"
/db_xref="GI:6940336"
/translation="MSNQSRHVRPAAEIVRELNRASPLRIDDENESWVNGYEKP
PTDLPVTSRDQLELLRSRHRIGLLVLPAPFAEQAVTGCFRVNVSGQELPDHRIAE
VLGICELDGKVEQIPTNVALRLRYDDLEQHEINDVSNLFTQEEFELWRDNCVNO
AISPPTHILTRKKVELYNALQLEAKPLSLIQTFTSFALRPOQKIGIMERHGVVYPWQ
LNHSLPFVVSQPTENPSRGKADTVDEEEDPGAAAPLLPKTEFN"
/complement(20231..28682)
/gene="FBgn0000155;roo"
/function="transposon"
/complement(20231..28682)
/gene="EG: BACR43E12.2"
/complement(41307..48656)
/gene="EG: BACR43E12.3"
/complement(41307..48656)
/gene="FBgn0002697;mgd1"
/function="transposon"
/complement(68339..69170)
/gene="EG: BACR43E12.7"
/complement(join(68339..69036,69092..69170))
/gene="EG: BACR43E12.7"
/note="/prediction=(method: 'genefinder',
version: '084', score: '34.12');
/prediction=(method: 'genscan', version: '1.0',
score: '98.21'); the EST embl|AA696429|AA696429 comes
from the 3' UTR"
/codon_start=1
/protein_id="CAB72280.1"
/db_xref="GI:6940337"
/translation="MQILHNFALLVLLAILANCDARRSAYRLPKLQGRSASGSPAP
QDVSGRDFEMKFMRLMGLFQQPPSDNVIVITQPSSTTTTPTGSPPTTTSTTTT
PTSNSTGRSLQESVHRQQPDDEAQETGVTTLTGRVLDRLDLPDGEEEPQELVTHK
HTQPKPKFNKMSVKLNNRKNLKKTHRRVHKHPQKKRKHKKLNKQSAANQQRNS
EPAVNFIEPYHQDRGQSQAGGQSTSDSRLQRIWRRFOIAN"
/complement(71879..72871)
/gene="EG: BACR43E12.6"
/complement(71879..72871)
/gene="EG: BACR43E12.6"
/note="/prediction=(method: 'genefinder',
version: '084', score: '38.91');
/prediction=(method: 'genscan', version: '1.0',
score: '99.03')
/codon_start=1
/protein_id="CAB72281.1"
/db_xref="GI:6940338"
/translation="MQPPPKTYNWLILLVLLGISLARTAPROATSTYVIPMSWNDLI
GTNFPEPQEKSHIEPEPRAFSVLDMSPDGQADDLQQLVMTSLGLTPAQIAQRSQ
GRNNEGRFFGRGGTTVVITGGAAANTNNNSFVSTPASTTTGTSTSEFIRTPIAYP
AGLPMQYFRKRQDDGTSMYSYPEFLRWNEAFYGGDLGSLGSNGAYGLSGVPLVPIVG
NEVRYVNMRLRMFRQLVGDPTTVASPPPSIREQDDQLEDDEDSALSIGPELVDES
PNEGQEVAGSGHSVADSPGFGILGRLRPLRRRPLQTLAQNIRRVQYVRK"
/complement(78912..79974)
/gene="EG: BACR43E12.5"
/complement(join(78912..79664,79735..79811,79896..79974))
/gene="EG: BACR43E12.5"
/note="/prediction=(method: 'genefinder',
version: '084', score: '49.15');
/prediction=(method: 'genscan', version: '1.0')
/codon_start=1
/protein_id="CAB72282.1"
/db_xref="GI:6940339"
/translation="MCGKGFMLVLIWATALSSDLSLVSGQVAPLTSSTSPAPKME
VKPTKPMVTGFTKSGNIYEIEDKRSIGSIEGRADQEQIVCNGVNIYSDVPCD
QKVNRYGEVKPLKDDRVEVTEKNHADGEQEQSQDMMQEQQDAEQDQEQDQD
HQVQSNHPRGQNNRRRRRPPQQQQQRLQQRRRRQQQKLQRRRRNGNGNNS
LRRRRNRNNNNLKRQQQRRRRPPGNNNNRRRLNNNNNNNNKQRQQHRRRQQQQ
QRRRRPNRNNRQQQQQRRRLHDDNI"
```

```
88303..90223
/gene="EG: BACR43E12.4"
join(88303..88527,89561..90223)
/gene="EG: BACR43E12.4"
/note="/match=(desc: 'ALPHA-DYSTROBREVIN-3',
species: 'Homo sapiens (Human)',
ranges: (query: 89621..89761,
target: SPTREMBL: O75697: 241..287, score: '105.00')),
method: 'blastx', version: '1.4.9');
/match=(desc: 'REF(2)p PROTEIN', species: 'Drosophila
melanogaster (Fruit fly)', ranges: (query: 89612..89713,
target: SWISS-PROT: P14199: 122..155, score: '107.00')),
method: 'blastx', version: '1.4.9'); the EST
embl|AI062288|AI062288 comes from the 5' UTR"
/codon_start=1
/protein_id="CAB72283.1"
/db_xref="GI:6940340"
/translation="MHSTSFRCGQOIRAAHAGRANIIIIIAAISIFGRLKGRCPAGK
TRKHCAFSPHRTYIVPGFPQCTPFALRLVSGSPSPMPRRQIGHTSGHCNVRCDGCG
NNRMTFYRYKCLHCLDYDLCSDCCKENGVSNGLSHLDPLQCLMDRDALELHEAGEPIP
ILCADSFTCPVCGEMGFVEDLTHCQDNHRMARTVCICPVCAAVPLSQPSHIAHIAN
HLMFSPSHRATGDPMDISATSQAEGLPQNSAVSSGSGAQHTFSSSENSQVRFLLP
GNRAPLADSDDFNMDNWEV"
BASE COUNT 28899 a 20617 c 21237 g 31868 t
ORIGIN

Query Match 11.5%; Score 45.4; DB 3; Length 102621;
Best Local Similarity 51.8%; Pred. No. 0.0086;
Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 49 aatagatgtgctaagatgtatgtttcgtgtgtgtattcttggttgcaagcaccaggcat 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83966 AACAGAAGGCCTTAATTATATGTGCGAGTAGTTGTTGTTGAGCTTAAGCAAGCCGTAT 83907

QY 109 gagtttggatgaatttaatacaagagagtaattcgttacaagaacttggttctgtgca 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83906 AAATATAAATTCAGCTAATGAATAACAATAAGCCGAAAATGCTAATGGAGTAGATCAAAA 83847

QY 169 acccaagaaaagtctgaaaaataaccaaatttagaaaaatgtaaggaatcagggggtttcaa 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83846 ATCAAAAAAAGATAAAACAATTAATAATAAGTAATAAATAAGGAATATGTA 83787

QY 229 gactctagatagcaagaag 247
||| ||| ||| ||| |||
Db 83786 TATACCAGATGGCGACGAG 83768

RESULT 8
AC104148/c
LOCUS AC104148 181443 bp DNA linear INV 27-DEC-2001
DEFINITION Drosophila melanogaster X BAC RP98-34O3 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.
ACCESSION AC104148
VERSION AC104148.3 GI:17981668
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 181443)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferreira,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
```


Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 300205)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

FEATURES
source
Location/Qualifiers
1. 300205
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
join(<12310..12326,13904..14602,14696..15533,24623..24826,24897..24980,25844..25954,26032..26269,26353..26546,27639..27828,27900..28343,29322..29779)
/gene="sgg"
/product="CT8875"
/db_xref="FLYBASE:FBan0002621"
/db_xref="FLYBASE:FBgn0003371"
<12310..>29779
/gene="sgg"
/note="CG2621"
/map="3B1-3B4"
/db_xref="FLYBASE:FBan0002621"
/db_xref="FLYBASE:FBgn0003371"
join(12310..12326,13904..14602,14696..15533,24623..24826,24897..24980,25844..25954,26032..26269,26353..26546,27639..27828,27900..28343,29322..29329)
/gene="sgg"
/note="sgg gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0002621"
/db_xref="FLYBASE:FBgn0003371"
/protein_id="AAF45801.1"
/db_xref="GI:7290342"

/translation="MFTFYTNINNTLINNNNNNTSNSSNNNNNVISOPIKIPLTER
FSSOTSGSADSGVIVSSASQQQLQPPPPRSSSSLSLPQAOPGGKWKQKQORQQLLL
SQDSGIENGVTTRPSKADNQAGKASHNATSSKESGASQSSSESLSGNCSEAQEQ
RVASSALESSVDTPVIVGGVSGNSILRSRIKYKSTNSTGTQEDVEDRIDEDVDI
CDDDDVCDRGSEIEEEEEDQTEEEVEDEVDKPKNRLPPDQAEALTVAAMARRR
DAKSLATDGHIFYFLLLKISEDPHIDSKLINRKDGLQDTMYYLDEFSGPKLREKFARKO
KQLLAKQOKQLMKRERSEQRKKNRTVASLNASGAVDDTKDKYKQPHCDTSSR
SKNSVPNPPSSHLHQNHNHLVDVQEDVDVNVATSDVDSGVVKNRRHSHDNHYDR
IPRSNAATITTRPIDQQSHHQNTEDVEQAEPIQIDGEADLDADADSDGSGENVK
TAKLARTQCKNQTGRDGSKIITVVATPGQTDVQEVSYTDTKVIENGSGFVVFQAK
LCDTGLVAIAIKKVLQDRFRKNREIQIMRKLEHCNIVKLLIYFFSYSGEKRDEVLNLVL
EYIPEYVKVARQYAKTKQTIPIPIRLMYQLFRSLAYIHSLGICHRDIKPQNLLLD
PETAVLKICDFGSAKOLLHGEPNVSYICSRYRAPELIFGAINYTKIDVWSAGCVLA
ELLGQPIPPGDSGVDQVLEVIKVLGTPTREQIREMNPNTYEFKFPQIKSHPWQKVF
IRTPTEAINLVSLLEYTPSARITPLKACAHPPFDELRMEGNHTLPNGRDMPPPLNF
EHELSPQSLVPQLLPHLQNASGPGNRPAGGAASIAASGSTSVSSSTSGSASVEGS
AQPSQGTAAAGSGGGATAGTGASAGGPGSGNNSSSGGASGAPSAVAAGGANAAY
AGGAGGGGAGATAAATATGAIGATNAGGANVTDs"
complement(join(35010..35971,36069..36458))
/gene="HLH3B"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT8939"
/db_xref="FLYBASE:FBan0002655"
/db_xref="FLYBASE:FBgn0011276"
complement(<35010..>36458)
/gene="HLH3B"
/note="CG2655"
/map="3B3-3B4"
/db_xref="FLYBASE:FBan0002655"
/db_xref="FLYBASE:FBgn0011276"
complement(join(35166..35971,36069..36393))
/gene="HLH3B"
/note="HLH3B gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0002655"
/db_xref="FLYBASE:FBgn0011276"
/protein_id="AAF45802.1"
/db_xref="GI:7290343"
/translation="MAWLPSSSNGADNADERSTASSGSGHSQTNESPRSGHLNGNGS
MRDTAAGRHPHALLRHATPHLPKTESISDGDGAELSDFLNTEDEEDLRYIVL
NGNQADANRSLSSSPRSHRNGLLTAPASSGSSVGGGGGNGSGGNASSGGSGVG
ATGVRKVFNTNTRERWRQNVSGAFELRKLVTPHPDKLKSNEILRSAIKYIKLLT
GILEWQORQAPSHPIRAQMEPNNDNRMANGHAADGENLENPDVPRHIKCERTDGO
MHRNGIGHGANGNAGNDLLMIAPGAVKSELLLESTLPLGLHPLNGPPLPTTAPLAM
AETQISGTVSGVKSASGRSSKRRLKPEGGATDSLIGKRRRT"
complement(36852..>37664)
/gene="EG:155E2.5"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT8935"
/db_xref="FLYBASE:FBan0002652"
/db_xref="FLYBASE:FBgn0025838"
complement(<36852..>37664)
/gene="EG:155E2.5"
/note="CG2652"
/map="3B3-3B4"
/db_xref="FLYBASE:FBan0002652"
/db_xref="FLYBASE:FBgn0025838"
complement(36852..37559)
/gene="EG:155E2.5"
/note="EG:155E2.5 gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0002652"
/db_xref="FLYBASE:FBgn0025838"
/protein_id="AAF45803.1"
/db_xref="GI:7290344"
/translation="MSYTNLRLHQNVVDDCQRVSCNRTPTTIVTGRCVISRDIVIGCR
MEQCDPMPYMLEPTWGVILRFRDGGDLRSFRVRRVTFKMSPRLPRLHLHVADSAPFEI

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Summary Statistics			

Contig length:	113559		
Phrap values in estimate:	110137		
Average error rate (BCM-Phrap estimate):	7.10641e-06		
Fraction of Phrap values less than 40 :	0.00837139		
Number of consensus changing edits:	464		
Number of N's in consensus :	0		

Consensus changing edits			

Position	Original+Context	Edited+Context	
16488	catctctcta(n)caatgatgac	catctctcta(g)caatgatgac	
32441	ccgaggcaac(n)tagagagacc	ccgaggcaac(a)tagagagacc	
33412	gagacatgaa(n)aatgtgtcgg	gagacatgaa(g)aatgtgtcgg	
46798	attgaacctc(n)atgcggttaag	attgaacctc(a)atgcggttaag	
50131	atagagaatg(n)tgagttaatg	atagagaatg(t)tgagttaatg	
50158	agggtacta(n)tttttcagat	agggtacta(c)tttttcagat	
50357	ctaataataa(n)ctggtcttat	ctaataataa(a)ctggtcttat	
52629	attctgatcc(c)ctcccacaga	attctgatcc(t)ctcccacaga	
52654	tcctggagg(g)ataacatcac	tcctggagg(g)tataacatcac	
53121	tatatattct(n)catctctctt	tatatattct(a)catctctctt	
53628	atataatat(n)ntatatata	atataatat(g)actatatata	
53629	tatatatatn(n)ntatatatan	tatatatatg(a)ctatatatac	
53630	atataatatn(n)tatatatana	atataatat(c)tatatatatac	
53639	ntatatata(n)agtcctatac	actatatata(c)agtcctatac	
55323	ccctccctg(n)nnccnnnnnt	ccctccctg(t)gtccatgtgt	
55324	ccctccctgn(n)ncnnnnntt	ccctccctgt(g)tccatgtgtt	
55325	ccctccctgn(n)cnnnnnntn	ccctccctgt(g)ccatgtgttc	
55328	ccctgnncc(n)nnnnntnnn	ccctgtgtcc(a)tggtgtctca	
55329	ccctgnnccn(n)nnntnnntn	ccctgtgtcca(t)gtgtgtctca	
55330	ctgnnnccnn(n)nnntnnntt	ctgtgtccat(g)tggttccatt	
55331	tnnnccnnn(n)ntnnnnntn	tggtgtccatg(t)gttctcattg	
55332	gnnnccnnn(n)tnnnnnntt	gtgtccatgt(g)ttctcattgt	
55335	nnccnnnnnt(n)nnnttntna	tccatgtgtt(c)tccattgttca	
55336	ccnnnnnttn(n)nnntnttnaa	ccatgtgttc(t)catgttcaa	
55337	cnnnnnnttn(n)ntnttnaan	catgtgttct(c)attgttcaac	
55338	nnnnnttnnn(n)tnnttnaan	atgtgttctc(a)ttgttcaact	
55341	nnntnnntt(n)tnnaannccc	tggttctcatt(g)ttcaactccc	
55344	tnnnntntt(n)aanncccnnt	tctcattgtt(c)aactcccact	
55347	ntntnttnaa(n)nnccnnntnn	catgtgttcaa(c)tcccacttat	
55348	ntntntnaah(n)ccnnnttnnn	attgttcaac(t)cccacttatg	
55352	tnnaannccc(n)ntnnnnnnn	ttcaactccc(a)cttatgagt	
55353	tnnaannccn(n)tnnnnnnnn	tcaactccca(c)ttatgagtga	
55356	annccnnnt(n)nnnnnnnaa	actcccactt(a)tgagtgaaga	
55357	nnccnnnttn(n)nnnnnnnaa	ctcccactta(t)gagtgaaga	
55358	nccnnnttnn(n)nnnnnnnaa	tcccacttat(g)agtgaagaata	
55359	ccnnnttnnn(n)nnnnnaann	ccacttatg(a)gtgaagaatat	
55360	ccnnnttnnn(n)nnnnnaann	ccacttatga(g)tgagaatatg	
55361	cnnttnnnnn(n)nnnaannng	cacttatgag(t)gagaatatgg	
55362	nnntnnnnnn(n)nnnaannng	acttatgagt(g)agaatatggg	
55363	ntnnnnnnnn(n)naannnggg	cttatgagt(g)aatatggggg	
55364	tnnnnnnnnn(n)aannngggg	ttatgagtga(g)aatatggggg	
55367	nnnnnnnaa(n)nnnggggnct	tgagtgaaga(t)atgggggtgct	
55368	nnnnnnnaa(n)nggggnnctt	gagtgaagaat(a)tggggtgctt	
55369	nnnnnnnaa(n)gggggncttg	agtgaagaata(t)gggggtcttg	
55374	naannngggg(n)ncctgggttt	gaatatggggg(t)gctgggtttt	
55375	aannnggggn(n)cttggttttc	aatatggggg(n)cttggttttc	
55619	agaatgattt(n)naancctttg	agaatgattt(a)taatcctttg	
55620	gaatgatttn(n)aanccttttg	gaatgattta(t)aatccttttg	

55623	tgatttnnaa(n)cctttgggta	tgatttnnaa(n)cctttgggta
58644	atccttagg(n)nngcagctac	atccttagg(c)tgagagctac
58645	tccttaggn(n)ngcagctaca	tccttaggc(t)ggagctaca
58646	ccttaggnn(n)gcagctacaa	ccttaggct(g)gcagctacaa
58754	acatacatit(n)ntntctcnt	acatacatit(c)attctctcat
58755	acatacattn(n)ttntctcnt	acatacatit(c)ttctctcat
58758	acatttnntt(n)ttctntnttt	acatttcatc(a)ttctctctt
58763	tnntntctc(n)tnntntntnn	tcattctctc(a)ttctgttcca
58765	ntntctcnt(n)ttntntnnnt	attctctcat(c)ttgtttcact
58769	tcctntnttt(n)tnnnntttt	tcctcatcttt(g)ttcacttttt
58772	cnntntntt(n)nnnttttttc	catcttgtt(c)actttttct
58773	ntntnttnn(n)nttttttcnn	atcttgtt(c)cttttttcta
58774	tnntnttnn(n)tttttttcnn	tccttgttca(c)ttttttctac
58782	nnntttttc(n)nnctngnacn	cacttttttc(t)actgggtaca
58783	nnnttttcn(n)ncnngnacna	acttttttct(a)ctgggtacaa
58784	ntttttcnn(n)ctngnacnac	cttttttcta(c)ctgggtacaa
58787	tttcnnnct(n)gnacnacctc	ttttctacct(g)gtacaacctc
58789	ttcnnnctng(n)acnacctctt	ttctacctgg(t)acaacctctt
58792	nnnctngnac(n)acnctctcnn	tacctgggtac(a)acctctcttg
58801	cnactcttc(n)nnnccncttc	caacctcttc(t)gtcccacttc
58802	naactcttcn(n)nnccnctctc	aaactcttct(g)ttccacttct
58803	acactctcnn(n)ncncttctn	acctcttctg(t)cccacttcta
58804	cttcctcnnn(n)ccnctctcnc	accttctgt(c)cccacttctac
58807	cttcnnnncc(n)cttctnccct	cttctgtccc(a)cttctacctt
58813	nnccncttct(n)ctttnccatn	tcccacttct(a)cttctcattta
58818	cttctnccct(n)catntttttc	cttctacctt(c)catatttttc
58823	nccttnccatt(n)tttctncttt	accttccatt(a)tttctcacttc
58829	catnttttct(n)ctttcttttt	cattattttc(a)cttctcttta

Query Match 10.9%; Score 43; DB 9; Length 113559;
Best Local Similarity 51.5%; Pred. No. 0.046;
Matches 123; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

QY	71	gtttcgggtgtgattctttgtgtgcaagcaccagcaggtgtttgtggaatttaataca	130
Db	49228	GTTACATTAATATTTTGGTTATATGCAACAGGGATCAACTCTGTCCACTTAAGTCCA	49287
QY	131	gagagtaattcgtttacaagaacttgggtttctgtgcaaccacgaagaaagctgaaaata	190
Db	49288	AAAAATGAATTGATTAGAGAAATGGATTACTCATCAAAATCAAAAGAAACCTGACA-A	49346
QY	191	acaaatatttagaaaatgtaaggaaatcaggggttttcaagactctagatagcaagaagtgg	250
Db	49347	CCAGGCTTTAGAAATGACAGAAACCAGGAAATTTTCAGTGATCTGGAGAGAACACCTAC	49406
QY	251	aaatcctgtctccagggaacctcctcaggggagaaatcagctccagctggtttatatcc	309
Db	49407	AGATGGTATAGTCAGGGAGACACTGAGATGATTATTTCAGTCCTAACCATTTTGTCTCC	49465

RESULT 13
AL359180/c
LOCUS Human DNA sequence from clone RP11-320G21 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL359180
VERSION AL359180.17 GI:12584420
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159229)
AUTHORS Tromans,A.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jan 28, 2001 this sequence version replaced gi:12329389.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/Chori
RP11-320G21 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-320G21. The true right end of clone RP11-94N9 is at 22847 in this sequence.

FEATURES

source

```
1. .159229
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-320G21"
/clone_lib="RPCI-11.2"
332.386
```

repeat_region	/note="LiMC2 repeat: matches 6269. .6325 of consensus"
repeat_region	422. .511
repeat_region	/note="LiM4 repeat: matches 5307. .5398 of consensus"
repeat_region	508. .688
repeat_region	/note="LiPA8 repeat: matches 4808. .4988 of consensus"
repeat_region	691. .1859
repeat_region	/note="LiPA8 repeat: matches 4990. .6155 of consensus"
repeat_region	2024. .2545
repeat_region	/note="LiME3A repeat: matches 5574. .6136 of consensus"
repeat_region	3110. .3537
repeat_region	/note="LiMA10 repeat: matches 5894. .6320 of consensus"
repeat_region	3899. .3958
repeat_region	/note="5 copies 12 mer 78% conserved"
repeat_region	3902. .3979
repeat_region	/note="13 copies 6 mer cttttc 66% conserved"
repeat_region	5497. .5949
repeat_region	/note="MLT1D repeat: matches 1. .505 of consensus"
repeat_region	6147. .6182
repeat_region	/note="3 copies 12 mer 94% conserved"
repeat_region	6150. .6179
repeat_region	/note="5 copies 6 mer ttattta 100% conserved"
repeat_region	7335. .7442
repeat_region	/note="MIR repeat: matches 36. .146 of consensus"
repeat_region	7802. .8018
repeat_region	/note="MIR repeat: matches 6. .240 of consensus"
repeat_region	8048. .8107
repeat_region	/note="10 copies 6 mer gagtag 73% conserved"
repeat_region	8051. .8104
repeat_region	/note="18 copies 3 mer gag 72% conserved"
repeat_region	10577. .10688
repeat_region	/note="MIR repeat: matches 151. .262 of consensus"
repeat_region	11719. .11778
repeat_region	/note="5 copies 12 mer 85% conserved"
repeat_region	11720. .11777
repeat_region	/note="29 copies 2 mer tc 82% conserved"
repeat_region	11781. .11816
repeat_region	/note="3 copies 12 mer 94% conserved"
repeat_region	12127. .12228
repeat_region	/note="MIR repeat: matches 50. .156 of consensus"
repeat_region	12732. .13280
repeat_region	/note="LiM1 repeat: matches 4580. .5151 of consensus"
repeat_region	13378. .13405
repeat_region	/note="14 copies 2 mer aa 89% conserved"

numbers given in the feature table with their source databases:
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP3-37005 is from the
 library RPCI-3 constructed by the group of Pieter de Jong. For
 further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 IMPORTANT: This sequence is not the entire insert of clone
 RP3-37005 It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP3-37005 is at 1 in this sequence. The
 right end of clone RP3-35AN1915 at 113155 in this sequence.

FEATURES source

source	Location/Qualifiers
	1. .113254
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP3-37005"
	/clone_lib="RPCI-3"
repeat_region	59. .142
	/note="42 copies 2 mer aa 63% conserved"
repeat_region	13347. .13438
	/note="HY4 repeat: matches 1. .93 of consensus"
repeat_region	14409. .14462
	/note="27 copies 2 mer tt 75% conserved"
repeat_region	17232. .17290
	/note="L1MA5A repeat: matches 5981. .6038 of consensus"
repeat_region	17291. .17382
	/note="L1MA5A repeat: matches 6199. .6294 of consensus"
repeat_region	17385. .17497
	/note="AluJ/monomer repeat: matches 1. .49 of consensus"
repeat_region	22889. .22973
	/note="MER82 repeat: matches 570. .653 of consensus"
repeat_region	22978. .23169
	/note="MER82 repeat: matches 83. .277 of consensus"
misc_feature	33394. .34525
	/note="CpG island"
	/evidence=not_experimental
repeat_region	36771. .37360
	/note="LTR1 repeat: matches 203. .785 of consensus"
repeat_region	37393. .37617
	/note="LTR1 repeat: matches 32. .258 of consensus"
repeat_region	37769. .39375
	/note="THE1C-internal repeat: matches 3. .1580 of consensus"
repeat_region	39378. .39755
	/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region	39760. .41138
	/note="MSTA-internal repeat: matches 1. .1442 of consensus"
repeat_region	46143. .46588
	/note="HERVL repeat: matches 4933. .5369 of consensus"
repeat_region	46867. .47592
	/note="HERVL repeat: matches 3966. .4698 of consensus"
repeat_region	48033. .48515
	/note="HERVL repeat: matches 1521. .2005 of consensus"
repeat_region	48756. .48901
	/note="HERVL16 repeat: matches 949. .1123 of consensus"
repeat_region	49431. .49550
	/note="60 copies 2 mer ct 60% conserved"
repeat_region	53032. .53393
	/note="MER39 repeat: matches 294. .688 of consensus"

		/note="L2 copies 2 mer ac 95% conserved"	
repeat_region	33008..33372		
	/note="MLTIAL repeat: matches 1..361 of consensus"		
repeat_region	33786..33842		
	/note="L2 repeat: matches 2688..2744 of consensus"		
repeat_region	33788..33856		
	/note="MIR repeat: matches 193..259 of consensus"		
repeat_region	35886..35919		
	/note="17 copies 2 mer aa 82% conserved"		
repeat_region	36181..36480		
	/note="AluX repeat: matches 1..302 of consensus"		
repeat_region	38591..38839		
	/note="L1M4 repeat: matches 2258..2507 of consensus"		
repeat_region	38840..39118		
	/note="AluSc repeat: matches 1..284 of consensus"		
repeat_region	39119..39182		
	/note="L1M4 repeat: matches 2507..2570 of consensus"		
repeat_region	39199..39874		
	/note="L1MA9 repeat: matches 5266..5928 of consensus"		
repeat_region	39878..39925		
	/note="4 copies 12 mer 83% conserved"		
repeat_region	39995..40281		
	/note="AluJo repeat: matches 1..284 of consensus"		
repeat_region	40282..41092		
	/note="L1MA9 repeat: matches 4495..5276 of consensus"		
repeat_region	41094..41568		
	/note="L1 repeat: matches 3294..3774 of consensus"		
repeat_region	41583..41813		
	/note="L1 repeat: matches 2911..3139 of consensus"		
repeat_region	41887..42845		
	/note="L1 repeat: matches 1940..2905 of consensus"		
repeat_region	42850..43085		
	/note="L1MC/D repeat: matches 5277..5507 of consensus"		
Query Match	10.8%;	Score 42.8;	DB 9; Length 159229;
Best Local Similarity	60.2%;	Pred. No. 0.053;	
Matches 71;	Conservative 0;	Mismatches 47;	Indels 0; Gaps 0;
QY	70	tgtttcgggtgattcttggttgcaagcaccaggcatgagtgttggtgaattaatca	129
Db	36599	TTTCTGTGCTAGATTCTTTGGCTGTAAAGCAACAGATGGGTTCTAATTAATTTAAACA	36540
QY	130	agagagtaattcggttacagaacttgggggtttctgtgcaacccaagaaaagtctgaaa	187
bh	26530	AGCACCGCAGAGGGAATAAGAAATGGCTAGGCATCTCTAAAGAAAGGCTAAA	36482

```
Query Match      10.8%; Score 42.8; DB 9; Length 159229;
Best Local Similarity 60.2%; Pred. No. 0.053;
Matches 71; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

[illegible]

RESULT 14	
AL137009	
LOCUS	AL137009 113254 bp DNA linear PRI 22-NOV-2001
DEFINITION	Human DNA sequence from clone Rp3-37005 on chromosome 6, complete sequence.
ACCESSION	AL137009
VERSION	AL137009.8 GI:9581553
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 113254)
TITLE	Bates, K.
JOURNAL	Direct Submission
COMMENT	Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 28, 2000 this sequence version replaced gi:9407728. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession


```
repeat_region 53394..53479
/note="LTR39 repeat: matches 711..794 of consensus"
repeat_region 53837..54314
/note="LTR39 repeat: matches 217..711 of consensus"
repeat_region 54616..54742
/note="MER21B repeat: matches 295..450 of consensus"
repeat_region 54742..54964
/note="LTR39 repeat: matches 1..226 of consensus"
repeat_region 63591..63618
/note="14 copies 2 mer tg 92% conserved"
repeat_region 63766..63813
/note="L1PA15-16 repeat: matches -209..-158 of consensus"
repeat_region 64934..64969
/note="18 copies 2 mer tc 91% conserved"
repeat_region 71439..71486
/note="24 copies 2 mer ca 97% conserved"
repeat_region 73918..75001
/note="MER11C repeat: matches 1..1071 of consensus"
repeat_region 77802..78110
/note="MER72 repeat: matches 382..668 of consensus"
repeat_region 78422..78555
/note="MER72 repeat: matches 244..382 of consensus"
repeat_region 79066..79195
/note="MER49 repeat: matches 2..131 of consensus"
repeat_region 81935..82045
/note="LTR28 repeat: matches 848..957 of consensus"
repeat_region 82035..82688
/note="LTR1 repeat: matches 1..662 of consensus"
repeat_region 82689..82782
/note="MLT2E repeat: matches 44..137 of consensus"
repeat_region 83634..84159
/note="LTR38 repeat: matches 1..547 of consensus"
repeat_region 92833..92932
/note="MLT1A2 repeat: matches 68..170 of consensus"
repeat_region 92937..93048
/note="MLT1A2 repeat: matches 1..115 of consensus"
repeat_region 100454..100649
/note="LTR33 repeat: matches 1..219 of consensus"
repeat_region 101865..101906
/note="21 copies 2 mer gt 100% conserved"
repeat_region 102970..103521
/note="MER89 repeat: matches 1..557 of consensus"
repeat_region 111796..111962
/note="L1PA5 repeat: matches 5979..6145 of consensus"
BASE COUNT 33925 a 23518 c 23177 g 32634 t
ORIGIN
Query Match 10.8%; Score 42.6; DB 9; Length 113254;
Best Local Similarity 49.8%; Pred. NO. 0.061;
Matches 163; Conservative 0; Mismatches 159; Indels 5; Gaps 2;
QY 69 atgttcggtgtgtattcttggttgcaagcaccaggcatgatttgggtgaatttaac 128
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91957 ATGTGTCACCTGAGAGTTTGTGGTGAAGCAACCAAAATCCAAATCTGACTAGTCAAAC 92016
QY 129 aagagagtaattcgttacagaacttgggtttctgtgcaacccaagaagaagtctgaaaa 188
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92017 TAGAAAAGATTTTTGGAGATTATGGACTGTTTCATTAATTAAGAAAGCAACTGACCA 92076
QY 189 taaccaaattagaaaaatgaaggaatcaggggtttcaagactctagatagaagaagt 248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92077 -ACCAGGCTTTGGAAAGGTAGGAATTAGGCTAGTCCAGGAATCTAGATAATGGGACTT 92135
QY 249 ----gaaaatcctgtctccaggggaacctcctcaggggagagaatcagctccagctggttta 304
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92136 GTCGGGCAGTATTTGTTCTTAGATACTATCATGAAAAATAAACTCGCTTCAGTATTTTTC 92195
QY 305 tatccaaagtgcattctcccaacctccaggaagtgcagagttctttagtatgtcttggat 364
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 92196 AATTCTTGTGTCACCTGCTTAACCTTCAAATTCCTGCTACTGTCTGGCCTGGTTGGGC 92255
QY 365 octatgcccaactttgtgaccagggcaa 391
```

```
Db 92256 CACATGCCCGTTTCTTGGAAGAGAAA 92282
||||| ||| ||| ||| ||| |||
RESULT 15
HUAC002302 259894 bp DNA linear PRI 23-NOV-1999
LOCUS Homo sapiens Chromosome 16 BAC clone CIT987-SKA-345G4 -complete
DEFINITION genomic sequence, complete sequence.
ACCESSION AC002302
VERSION AC002302.1 GI:2576341
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 259894)
Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
Genome duplications and other features in 12 Mb of DNA sequence
from human chromosome 16p and 16q
Genomics 60 (3), 295-308 (1999)
99425270
10493829
2 (bases 1 to 259894)
Adams,M.D., Loftus,B.J., Zhou,L., LaBombard,M., Fuhrmann,J.,
Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
Homo sapiens Chromosome 16 BAC clone CIT987-SKA-345G4 #complete
genomic sequence
Unpublished
3 (bases 1 to 259894)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (19-JUN-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, Email:
mdadams@tigr.org
4 (bases 1 to 259894)
Adams,M.D.
Direct Submission
Submitted (30-OCT-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
5 (bases 1 to 259894)
Adams,M.D.
Direct Submission
Submitted (17-DEC-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Oct 30, 1997 this sequence version replaced gi:2570249.
Address all correspondence to:
Mark Adams
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA
e-mail address: mdadams@tigr.org. The bac location is on
chromosome BAC clone is located on human chromosome 16p12.2-p12.
The orientation of the sequence is from SP6 end to T7 end.
Genes were identified by a combination of five methods including:
XGRAIL (available by anonymous ftp from arthur.epm.ornl.gov),
Genefinder (Phil Green, University of Washington), Genscan (Chris
Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html
)
searches of the complete sequence against a peptide database,
and the Human gene Index database at TIGR
(http://www.tigr.org/tdb/hg
i/hgi.html).
A gene with homology to another protein is annotated as the isolog
of
that protein.
Genes without pepetide homolgy having spliced EST hits are termed
'u
```



```
QY 189 taacaaa-----tttagaaaaatgtaaggaaatcagggggtttcaagactcttagatagc 241
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124770 GATGAAGAACCAGGCTTTAGAAAAGTCCTGAAACCAAAATGAATTCTAGGGATCTAGGAAGC 124829

QY 242 aagaagtggaaaaatcctgtctccaggggaacctcctcaggggagagaatca 289
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124830 AGGGACTAATGACTGTTTCCTTCAGGACACCCAGAGATGGGGTGAATAA 124877
```

Search completed: August 26, 2002, 23:11:39
Job time: 6339 sec